

# **Suicides in India**

Venkata Sai Kumar, Rachagundla

( 11696627 )

Department of Information Science, University of North Texas

INFO 5709: Data Visualization and Communication

Dr. Gahangir Hossain

December 13, 2023

## **Introduction:**

Suicide, a significant public health issue, is often influenced by a multitude of factors including mental health, financial conditions, societal pressures, and more. Additionally, understanding the reasons behind these suicides - whether they are related to family problems, illness, financial distress, or other causes - can guide more effective and tailored prevention strategies. In a country as vast and varied as India, understanding the patterns and trends of suicides is essential for creating effective interventions and policies.

Visualizing this data using tools like Tableau can make the information more accessible and comprehensible to a wider audience. Interactive dashboards can facilitate a deeper understanding of the data, allowing for the identification of trends over time and correlations between different factors. This visual approach can be particularly effective in communicating complex data to non-specialist audiences, including policymakers and the general public.

In conclusion, the "Suicides in India 2001-2012" dataset serves as a vital resource for comprehending the multifaceted nature of suicides in India. Through exploratory data analysis and visualization, it can provide key insights into the patterns and causes of suicides, which is essential for formulating effective prevention strategies and policies. This analysis not only contributes to the field of mental health and social welfare but also supports broader efforts to address public health challenges in India.

## **Dataset Description:**

The dataset is collected from Kaggle. This dataset compiles historical data on suicides that occurred in India between the years 2001 and 2012. It encompasses a wide range of variables and dimensions, including information about the geographical locations of suicides, the motivations or reasons behind these tragic events, the number of individuals affected, and the underlying causes of the suicides that transpired throughout India during this 12-year period. It includes 2,37,519 rows of data with seven attributes: State, Year, Type code, Type, Gender, Age group, Total.

### ***Data Source:***

<https://www.kaggle.com/datasets/rajanand/suicides-in-india>

### **Attributes:**

Total of 7 attributes are used to describe the Suicides in India. Below is the list

1. **State:** Name of Indian State where Suicide happened.
2. **Year:** The year Suicide happened.
3. **Type Code:** It mainly shows the causes which is categorized as Social, Educational, Professional, Social Status.
4. **Type:** It further categorizes the Type code that is the causes.
5. **Gender:** Gender of the Suicided (Male/Female).
6. **Age Group:** Age group of the Suicided (0-14, 15-29, 30-44, 45-59, 60+, 0-100+).
7. **Total:** Total number of people committed Suicide.

### **Tools:**

- Python
- Tableau

### **Data Cleaning:**

- Some of the states in the States Column have abbreviations like UT for union territories. Removing those abbreviations for the state column.
- Removing some of the samples like Total (UT's), Total (All India) etc. in the States column.

```
✓ [146] # Renaming States
0s df.replace('Delhi (Ut)', 'Delhi', inplace=True)
```

```
✓ # Dropping unnecessary rows
0s df = df.drop(df[(df.State == 'Total (Uts)') | (df.State == 'Total (All India)') |
                 (df.State == 'Total (States)')].index)
```

## Exploratory Data Analysis:

### Importing modules and reading data:

```
✓ # Importing modules
0s import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

```
✓ [4] # Read data from CSV file
df = pd.read_csv('/content/Suicides in India 2001-2012.csv')
```

```
✓ [5] df.head()
```

	State	Year	Type_code	Type	Gender	Age_group	Total	
0	A & N Islands	2001	Causes	Illness (Aids/STD)	Female	0-14	0	
1	A & N Islands	2001	Causes	Bankruptcy or Sudden change in Economic	Female	0-14	0	
2	A & N Islands	2001	Causes	Cancellation/Non-Settlement of Marriage	Female	0-14	0	
3	A & N Islands	2001	Causes	Physical Abuse (Rape/Incest Etc.)	Female	0-14	0	
4	A & N Islands	2001	Causes	Dowry Dispute	Female	0-14	0	

### **Information about data:**

✓  
0s

```
[6] df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 237519 entries, 0 to 237518
Data columns (total 7 columns):
#   Column      Non-Null Count  Dtype  
---  -
0   State       237519 non-null object  
1   Year        237519 non-null int64   
2   Type_code   237519 non-null object  
3   Type        237519 non-null object  
4   Gender      237519 non-null object  
5   Age_group   237519 non-null object  
6   Total       237519 non-null int64   
dtypes: int64(2), object(5)
memory usage: 12.7+ MB
```

### **Null Values and Duplicates:**

✓  
0s

```
[10] df.isnull().sum()
```

```
State      0
Year       0
Type_code  0
Type       0
Gender     0
Age_group  0
Total      0
dtype: int64
```

✓  
0s

```
[11] df.shape
```

```
(237519, 7)
```

✓  
0s

```
[12] df.duplicated().sum()
```

```
0
```

Our data does not have any missing values and there are no duplicates in our dataset. So, there is no need to drop any rows.

### ***1. Which state is having a greater number of Suicides in India over the years?***

```
✓ 0s ▶ import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Grouping by 'State' and calculating the total number of suicides for each state
statewise_suicides = df.groupby('State')['Total'].sum()
statewise_suicides_df = statewise_suicides.reset_index()
statewise_suicides_df.rename(columns={'Total': 'Total Suicides'}, inplace=True)

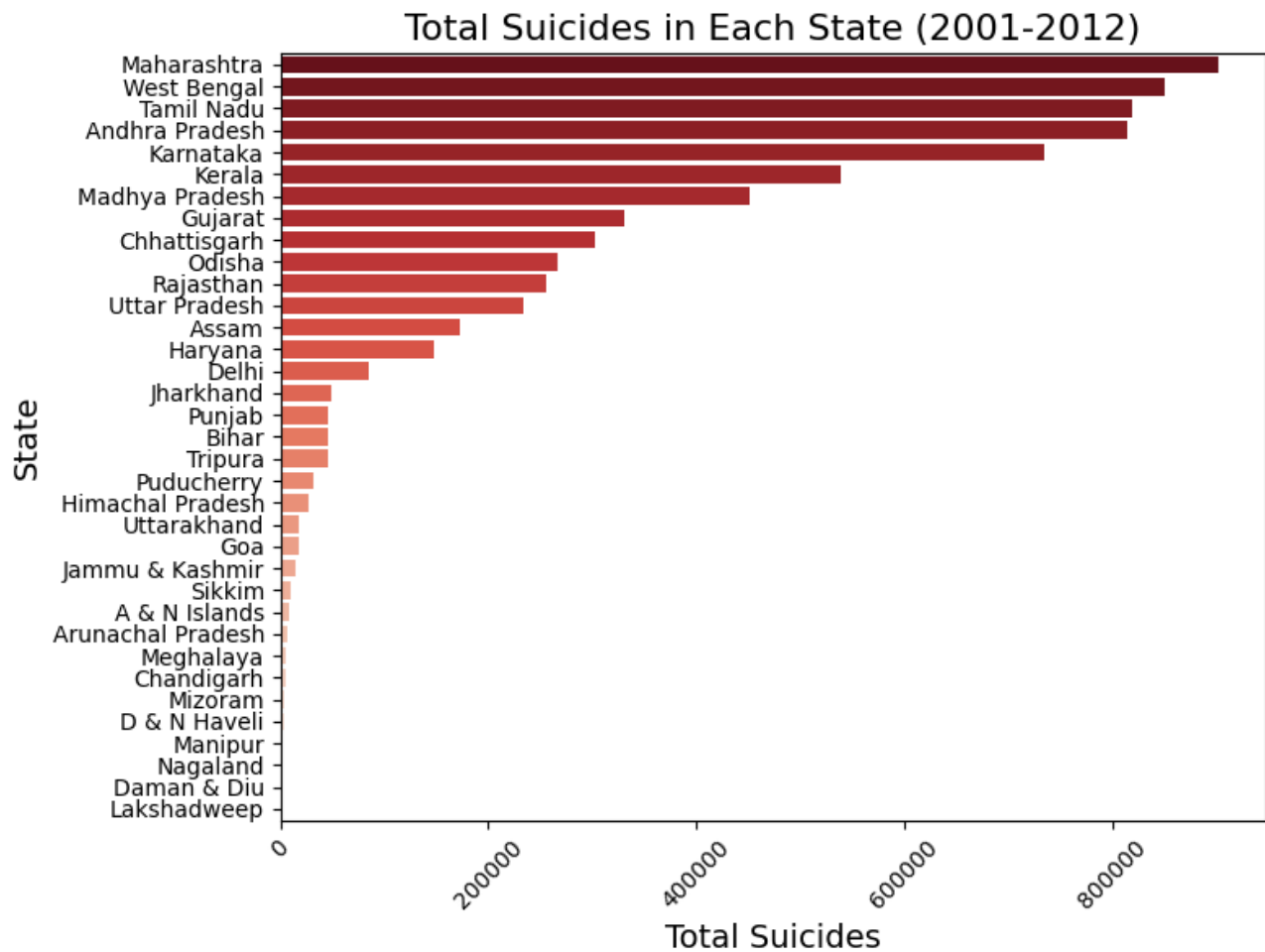
# Sorting the data in descending order of total suicides
sorted_suicides_df = statewise_suicides_df.sort_values(by='Total Suicides', ascending=False)

# Reversing the color palette from dark red to light red
reversed_palette = sns.color_palette("Reds_r", len(sorted_suicides_df))

# Creating the bar plot with the reversed color palette
plt.figure(figsize=(8, 6))
sns.barplot(x='Total Suicides', y='State', data=sorted_suicides_df, palette=reversed_palette)

plt.title('Total Suicides in Each State (2001-2012)', fontsize=16)
plt.xlabel('Total Suicides', fontsize=14)
plt.ylabel('State', fontsize=14)
plt.xticks(rotation=45)
plt.tight_layout()

# Show the plot
plt.show()
```



### **Results:**

Above visualization depicts that Maharashtra has the highest number of Suicides over the years in India whereas Lakshadweep is having the least number of Suicides over the years in India. We can also visualize that the number of Suicides in Union Territories are less compared to the states in India.

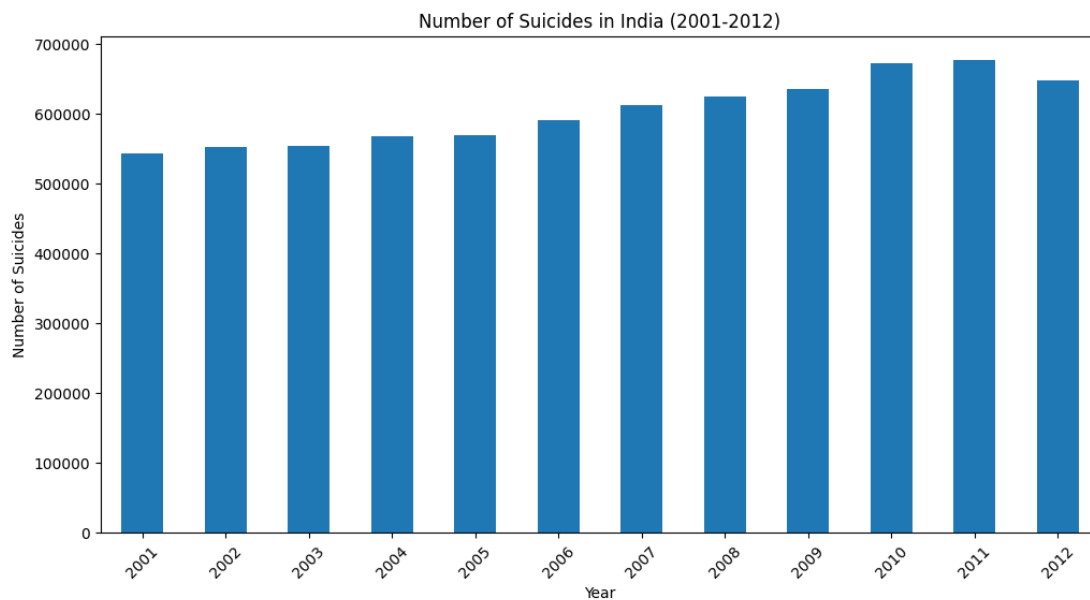
## 2. Which Year has the highest Suicides and how is the trend of suicides from 2001 to 2012?

✓  
0s

```
# Grouping the data by year and summing the total suicides
yearly_suicides = df.groupby('Year')['Total'].sum()

# Plotting the data
plt.figure(figsize=(12, 6))
yearly_suicides.plot(kind='bar')
plt.title('Number of Suicides in India (2001-2012)')
plt.xlabel('Year')
plt.ylabel('Number of Suicides')
plt.xticks(rotation=45)

# Show the plot
plt.show()
```



### **Results:**

Above visualization depicts that Year- 2011 has the highest number of Suicides in India. Overall, the trend is gradually increasing over time.



### 3. Which category of the gender committed a greater number of Suicides in India over the years?

```
0s # Grouping the data by 'Gender' and calculating the total number of suicides for each gender
genderwise_suicides = df.groupby('Gender')['Total'].sum()

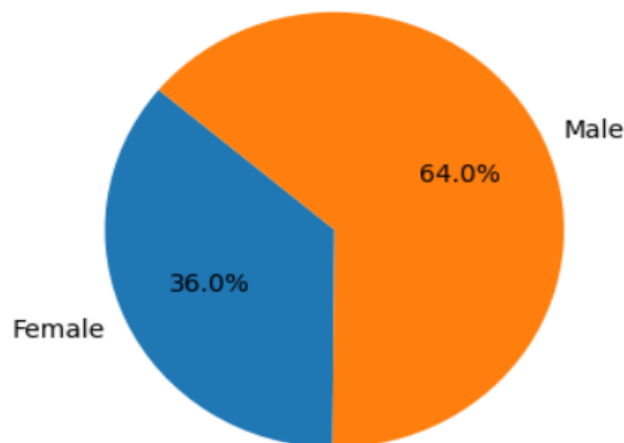
# Converting the result to a DataFrame for better readability
genderwise_suicides_df = genderwise_suicides.reset_index()
genderwise_suicides_df.rename(columns={'Total': 'Total Suicides'}, inplace=True)

# Creating the pie chart
plt.figure(figsize=(8, 8))
plt.pie(genderwise_suicides_df['Total Suicides'], labels=genderwise_suicides_df['Gender'], autopct='%1.1f%%', startangle=140)
plt.title('Gender Distribution of Suicides (2001-2012)')

# Show the plot
plt.show()

genderwise_suicides_df.head()
```

Gender Distribution of Suicides (2001-2012)



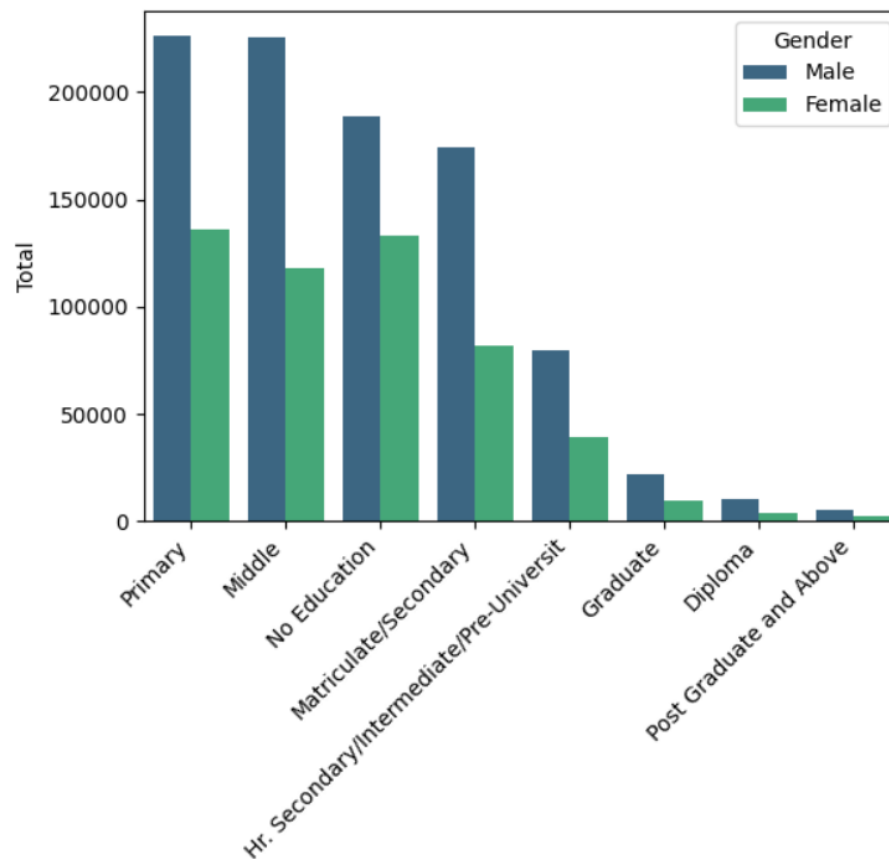
	Gender	Total Suicides
0	Female	2606922
1	Male	4641088

#### **Results:**

From the Output we can illustrate that males committed more number of Suicides than females in India.

**4. Which educational status category has the highest number of suicide cases associated with it?**

```
eduDf = df[df['Type_code']=='Education_Status']
plt.figure(figsize=(6,6))
eduDf = eduDf[['Type', 'Gender', 'Total']]
edSort = eduDf.groupby(['Type', 'Gender'], as_index=False).sum().sort_values('Total', ascending=False)
sns.barplot(x='Type', y='Total', hue='Gender', data=edSort, palette='viridis')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
```

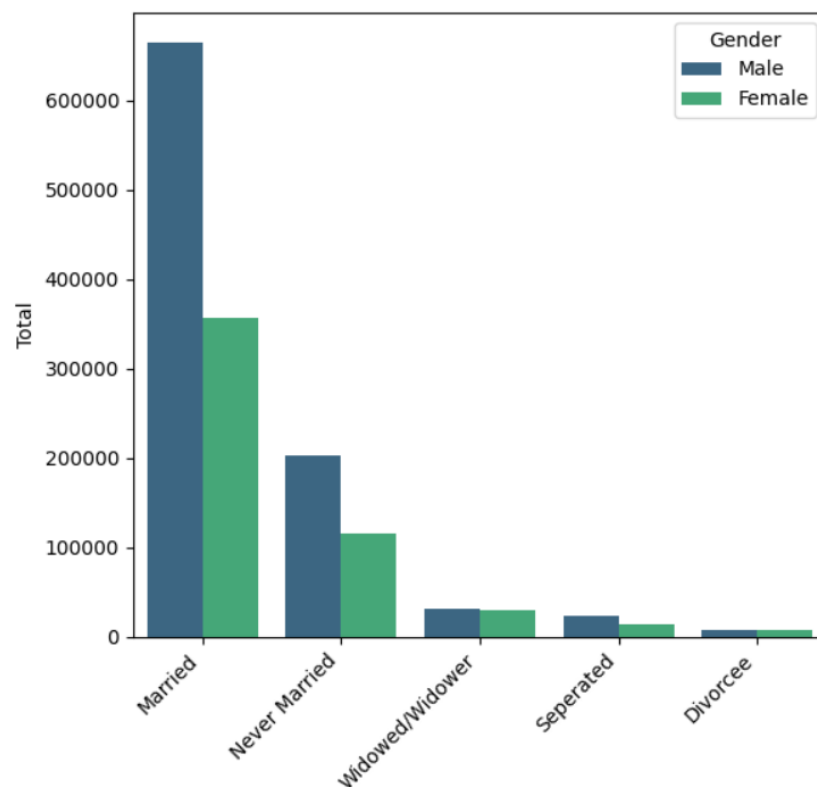


**Results:**

The highest number of Suicide cases are associated with the Primary school Educational Status and the lowest number of Suicide cases are associated with Postgraduate and above. We can also depict that more males are committing Suicide than females irrespective of Educational Status.

**5. Which Social status category has the highest number of Suicide cases associated with it?**

```
plt.figure(figsize=(6,6))
socialDf = df[df['Type_code']=='Social_Status']
socialDf = socialDf[['Type', 'Gender', 'Total']]
socialSort = socialDf.groupby(['Type', 'Gender'], as_index=False).sum().sort_values('Total', ascending=False)
sns.barplot(x='Type', y='Total', data=socialSort, hue='Gender', palette='viridis')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
```

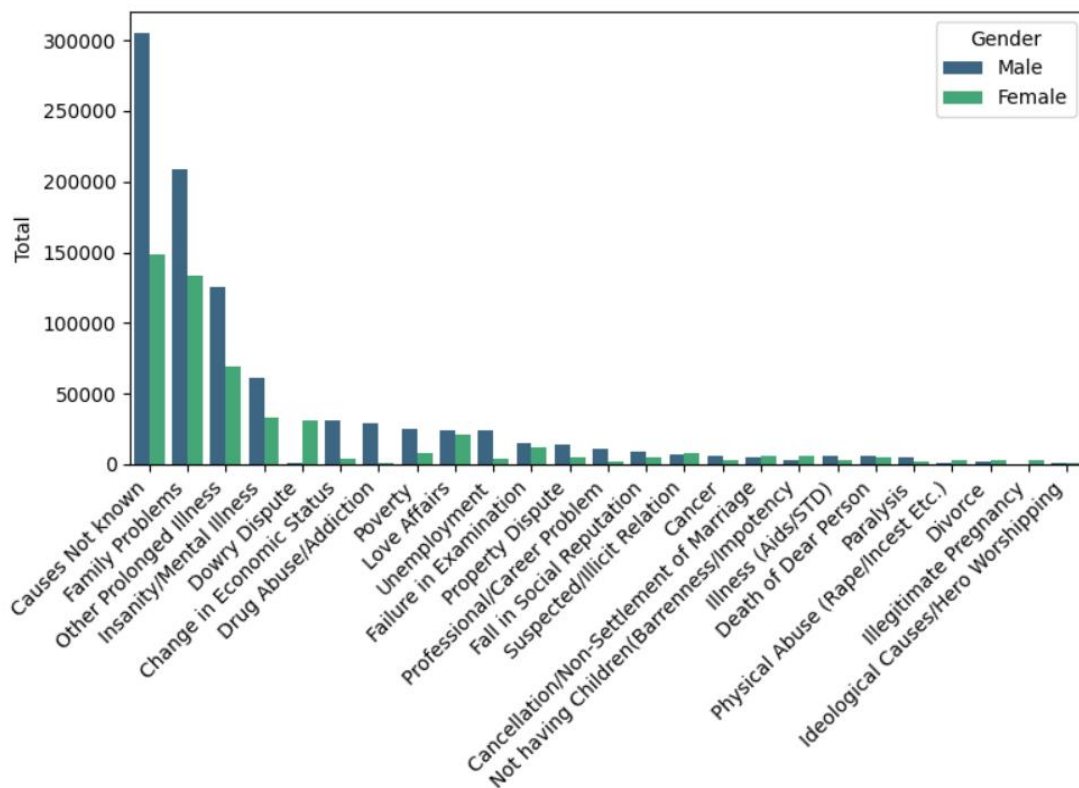


**Results:**

The highest number of Suicide cases are associated with the Married Status and the lowest number of Suicide cases are associated with Divorcee. We can also depict that more males are committing Suicide than females irrespective of Social Status.

## 6. What are the causes for Suicides in India?

```
causesDf = df[df['Type_code']=='Causes']
causesDf.loc[causesDf['Type']=='Bankruptcy or Sudden change in Economic','Type'] = 'Change in Economic Status'
causesDf.loc[causesDf['Type']=='Bankruptcy or Sudden change in Economic','Type'] = 'Change in Economic Status'
causesDf.loc[causesDf['Type']=='Other Causes (Please Specity)','Type'] = 'Causes Not known'
causesDf.loc[causesDf['Type']=='Not having Children (Barrenness/Impotency)','Type'] = 'Not having Children(Barrenness/Impotency)'
plt.figure(figsize=(8,6))
causesDf = causesDf[['Type','Gender','Total']]
causesSort = causesDf.groupby(['Type','Gender'],as_index=False).sum().sort_values('Total',ascending=False)
sns.barplot(x='Type',y='Total',data=causesSort,hue='Gender',palette='viridis')
plt.xticks(rotation=45,ha='right')
plt.tight_layout()
```



### Results:

For most of the Suicides in India causes are not known barring that Family problems are the main cause for most of the Suicides occurring in India. We can also depict that more males are committing Suicide than females in many cases irrespective of the causes.

### **Hypothesis:**

- Does the trend of suicides in India show significant variations when analyzed across different time periods, age groups, and states, indicating specific demographic and geographic factors influencing the suicide rates?
- Marital status significantly impacts the suicide rates across genders in India, with distinct patterns emerging over time and varying distributions of suicide causes among different marital categories.
- The prevalence of suicides in India is significantly associated with the professional and educational status of individuals, as well as the methods adopted for suicide, suggesting that socioeconomic factors may influence the likelihood and means of suicide.

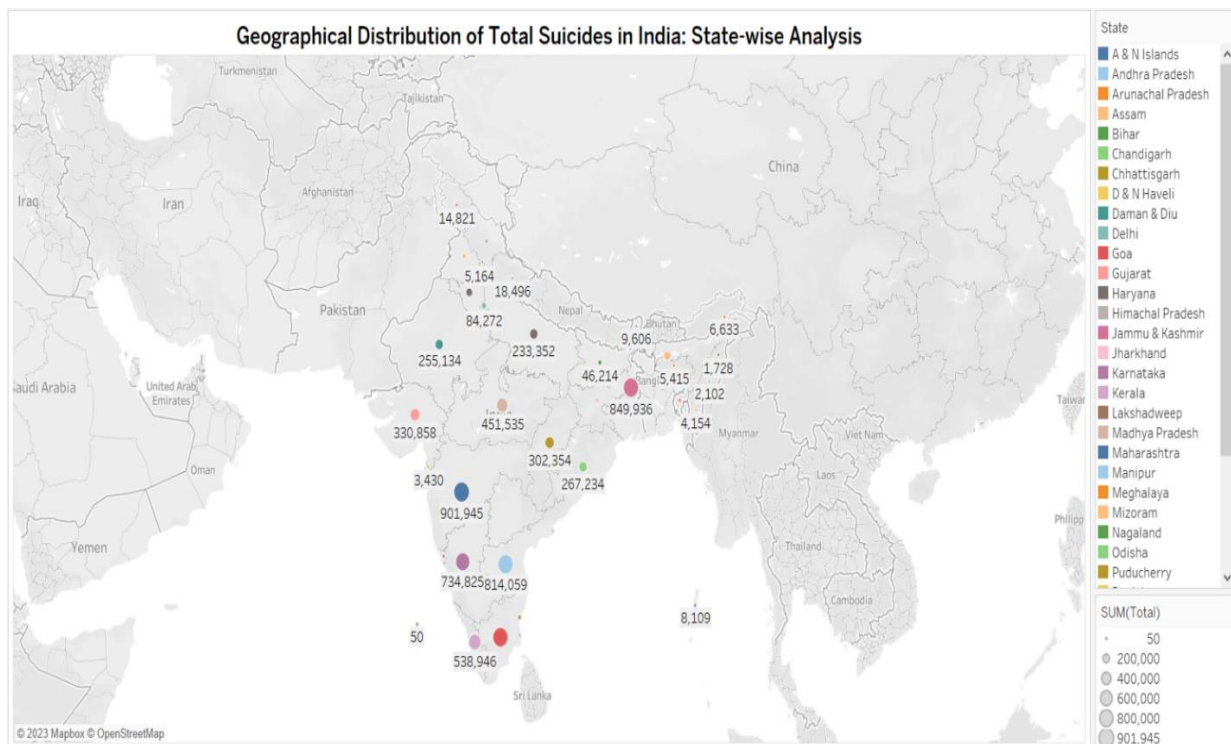
## Hypothesis-1

Does the trend of suicides in India show significant variations when analyzed across different time periods, age groups, and states, indicating specific demographic and geographic factors influencing the suicide rates?

## Maps

**Design Rationale:** The map visualization effectively utilizes geographic distribution to illustrate state-wise variations in the total number of suicides in India.

**Perceptual Tasks:** By employing a color-coded scheme, where colors represent States allows for immediate visual comparison across states. Size variation of the circles represent the total number of suicides, making it perceptually easy to identify States with larger Suicides. Labelled the total number of Suicides in a State.

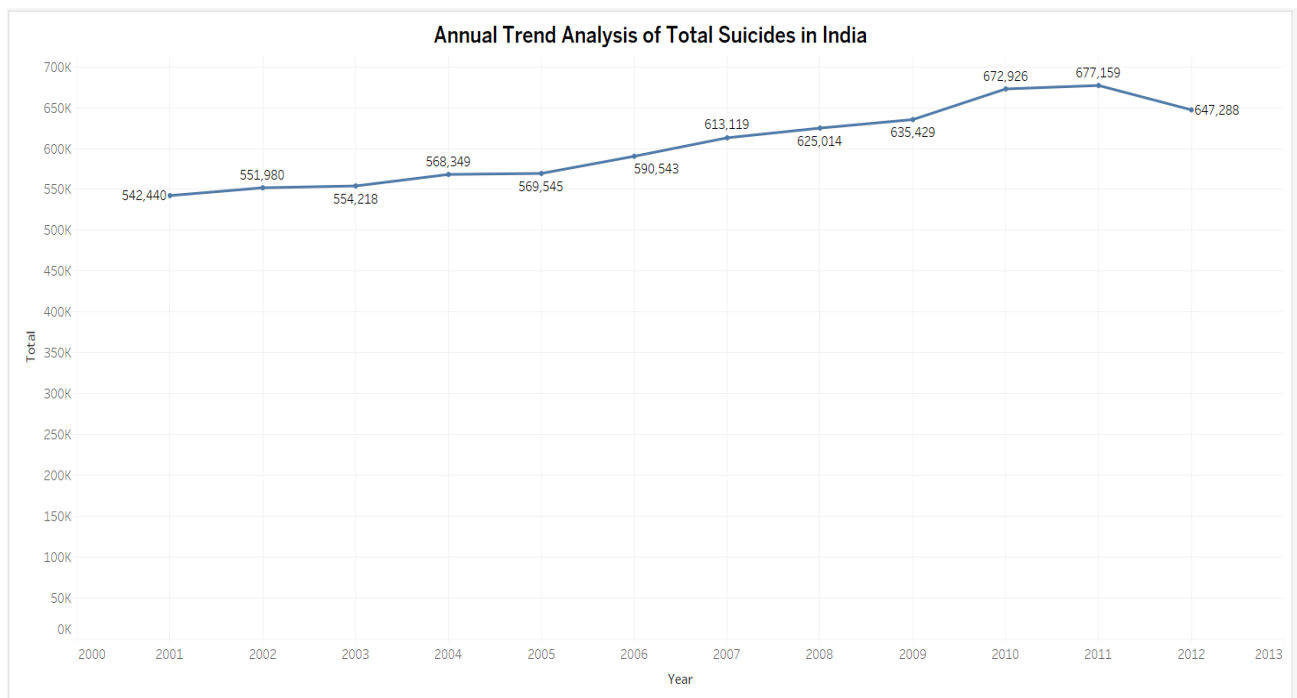


## Results:

From the above map we can identify which states have higher or lower suicide totals. Highest number of Suicides are reported in Maharashtra whereas the least number of Suicides are reported in Lakshadweep.

## Line graph:

**Design Rationale:** Line graph effectively illustrates the year-over-year trend analysis of a total suicides in India. The graph succinctly communicates changes over time, allowing for quick visual assessment of upward or downward trends in the data.



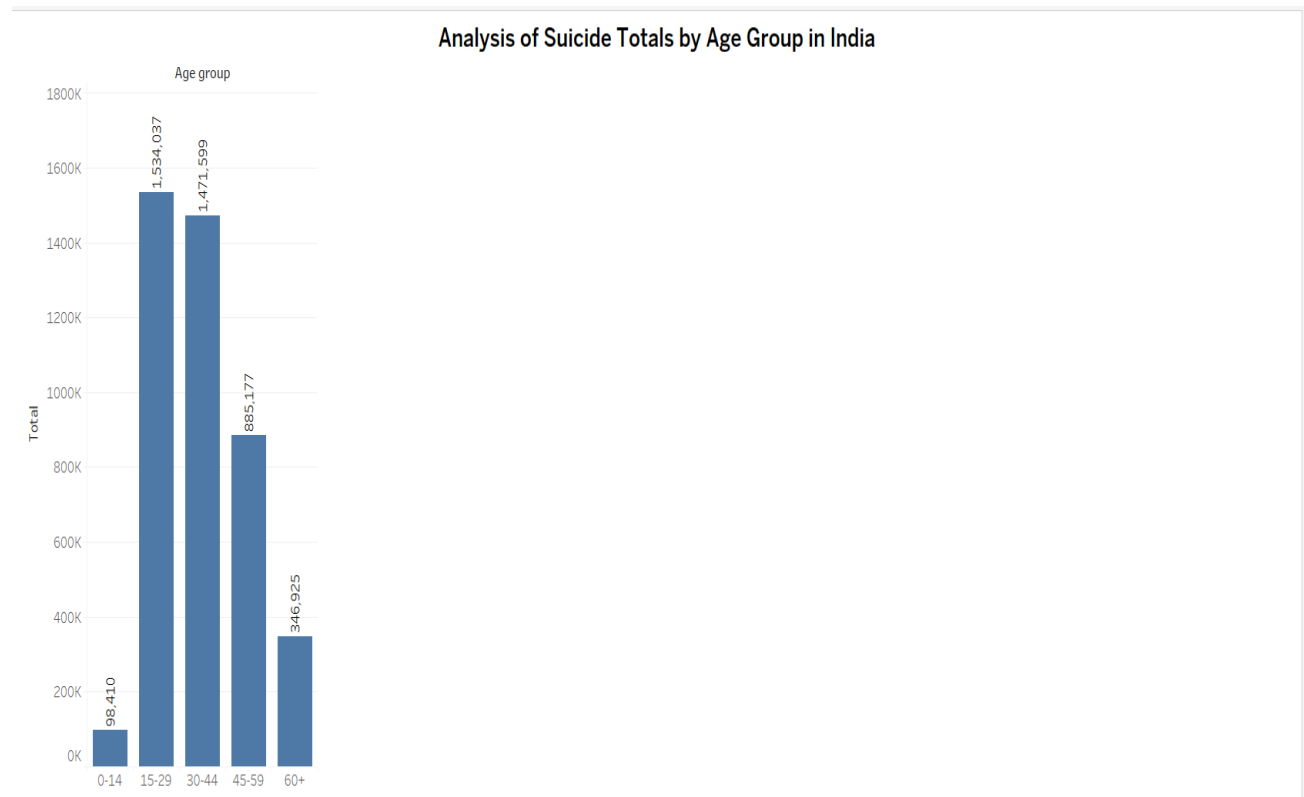
## Results:

The line chart displayed indicates a general upward trend in the total number of suicides in India from the year 2001 to 2012, with a slight decrease observed in the final year. The ascending line suggests that over the 12-year span, the occurrence of suicides has increased, with some fluctuations.

## **Bar graph:**

**Design Rationale:** The bar chart is a classic and effective visualization tool for comparing the Total number of suicides across different age demographics within India.

**Perceptual Tasks:** The length of the bars allows for quick assessment and comparison of the number of suicides across age groups.



## **Results:**

This bar chart indicates that the age groups 15-29 and 30-44 have the highest total number of suicides, with these age groups typically representing young adults to middle-aged individuals. The numbers appear to decrease for the subsequent older age groups and are the lowest for the youngest age group (0-14) and the oldest (60+).



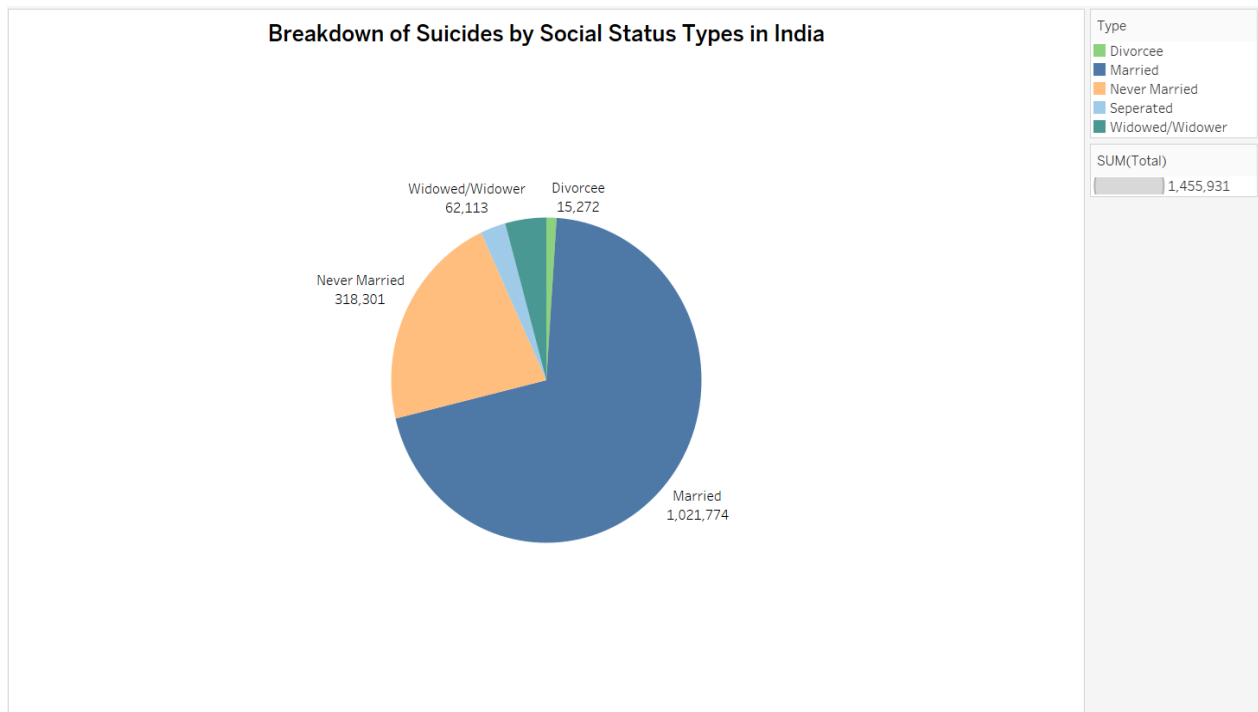
## Hypothesis-2

Marital status significantly impacts the suicide rates across genders in India, with distinct patterns emerging over time and varying distributions of suicide causes among different marital categories.

### Pie Charts:

**Design Rationale:** A pie chart has been chosen for this visualization to show the proportion of suicides by different social status types in India.

**Perceptual Tasks:** The different colors assigned to each segment facilitate distinction among the social status types. Size comparison is intuitive as the viewer can easily perceive which segments of the pie are larger, indicating a higher count of suicides.



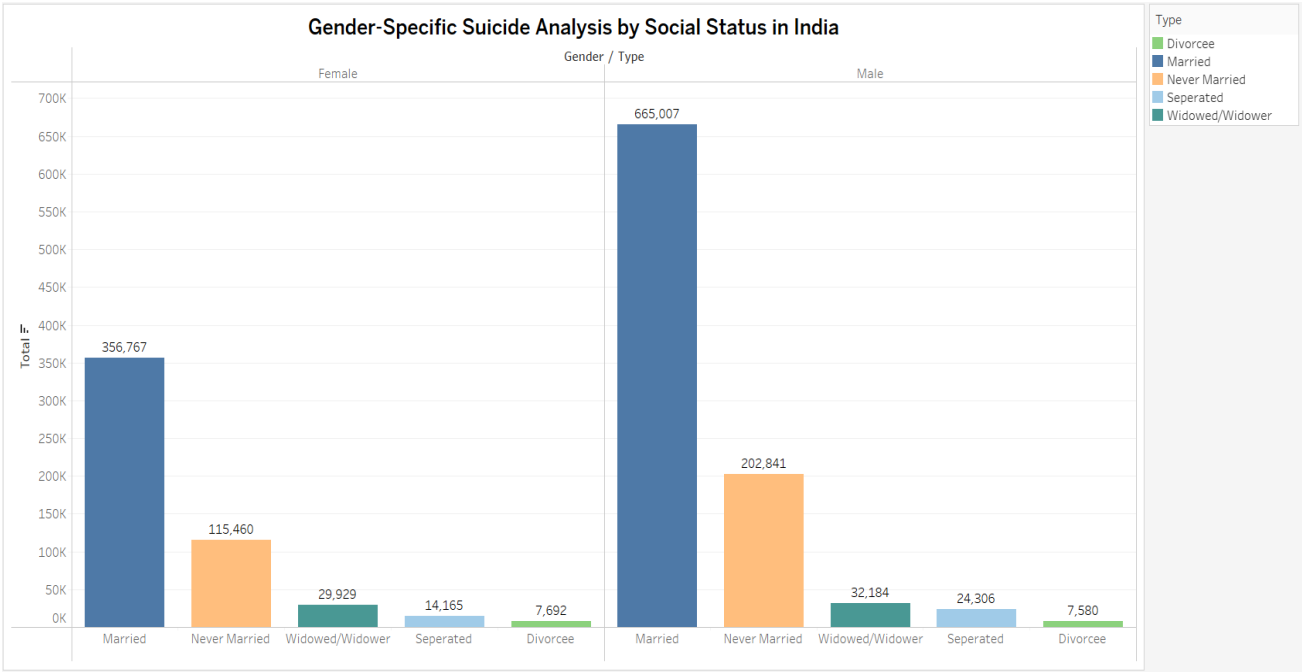
### Results:

The pie chart indicates that the majority of suicides are reported among married individuals, representing the largest segment of the chart. The 'Never Married' status forms the second-largest category, followed by 'Widowed/Widower', 'Separated', and 'Divorcee', respectively.

Side-by-Side bars

**Design Rationale:** The given visualization employs a side-by-side bar chart to compare the total number of suicides across different social statuses, segmented by gender.

**Perceptual Tasks:** The bars lengths provide a quick visual indication of the magnitude, facilitating immediate comparison between groups. Distinct colors represent different social statuses, aiding in distinguishing between them. The gender separation within each social status category allows for a comparison that highlights the gender disparity in suicide cases.



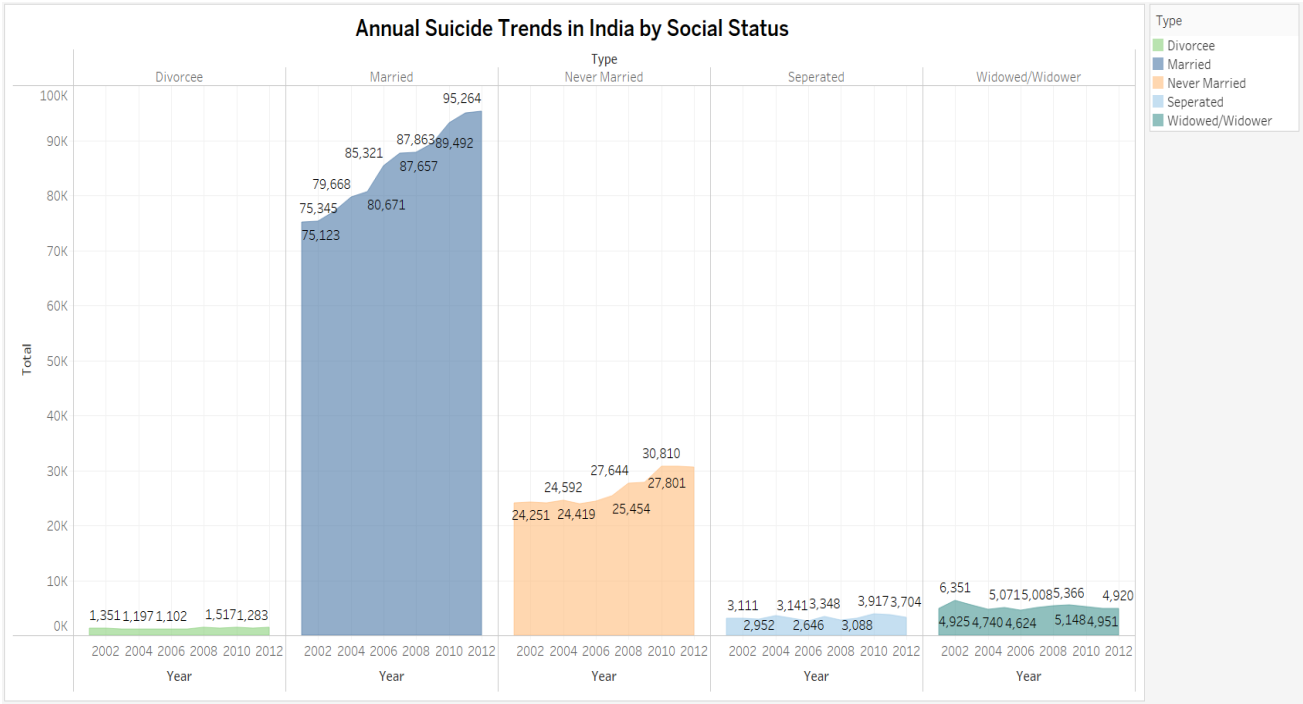
Results:

The visualization reveals that within the dataset, married individuals have the highest total number of suicides among both genders, with a notably higher incidence in males. Widowed, separated, and divorcee categories have significantly lower totals, with the lowest suicides among divorcees.

Area Charts

**Design Rationale:** The Area chart is employed to demonstrate the annual suicide trends in India by social status.

**Perceptual Tasks:**  
Colors differentiate between social status types, enabling the viewer to distinguish between categories such as married, never married, separated, and widowed/widower. The stacked layers of color in the area chart illustrate the proportion of each social status type relative to the total, highlighting trends and changes over time within each category.



Results:

The visualization reveals distinct patterns in suicide trends based on social status over the years. For example, it appears that the married category consistently has the highest number of suicides across all years. The divorce category has comparatively lower numbers. Additionally, the overall trend suggests an increase in the number of suicides over the years, with some fluctuation.

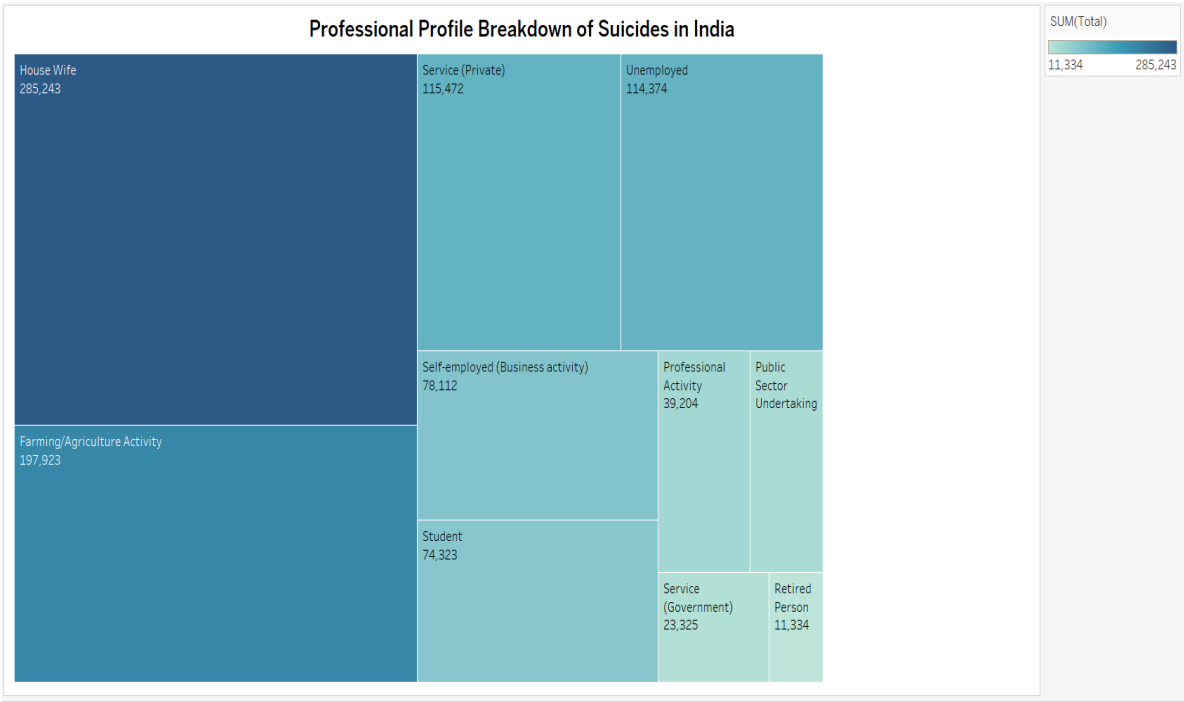
**Hypothesis-3**

The prevalence of suicides in India is significantly associated with the professional and educational status of individuals, as well as the methods adopted for suicide, suggesting that socioeconomic factors may influence the likelihood and means of suicide.

**Tree Maps**

**Design Rationale:** A tree map is chosen for its ability to display hierarchical data and part-to-whole relationships. Here, each rectangle represents a professional category, and its size indicates the number of suicides associated with that profession.

**Perceptual Tasks:** Viewers can quickly perceive which professional categories have the largest and smallest number of suicides based on the size of the rectangles. The darker color shades represent higher numbers of Suicides provides an immediate visual cue to identify which professional groups have higher or lower suicide counts.



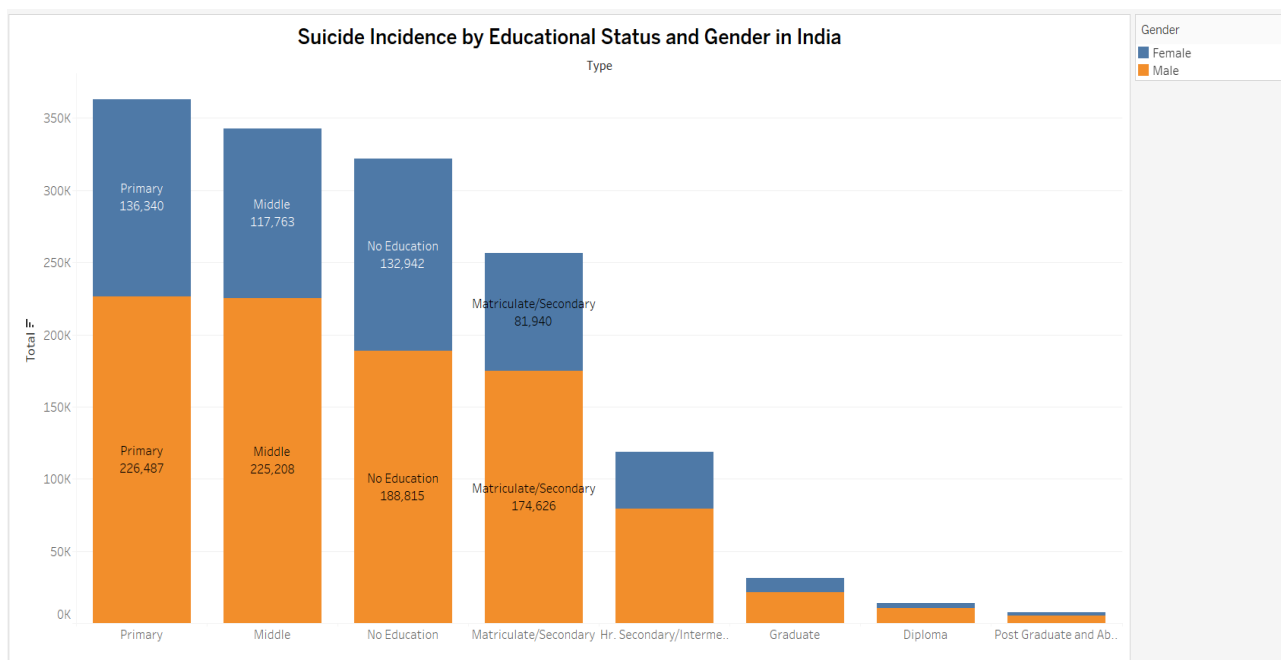
**Results:**

Housewife are committing a greater number of Suicides whereas Retired Persons are committing fewer number of Suicides.

## Stacked Bar Chart

**Design Rationale:** Stacked bar charts are suitable for displaying the distribution of different subgroups within a larger category. In this case, it helps to compare the total number of suicides across different educational statuses while breaking down each category by gender.

**Perceptual Tasks:** We can compare the total number of suicides across different educational categories and observe the gender distribution within each category. The color-coded segments effectively highlight gender-specific trends within the data.



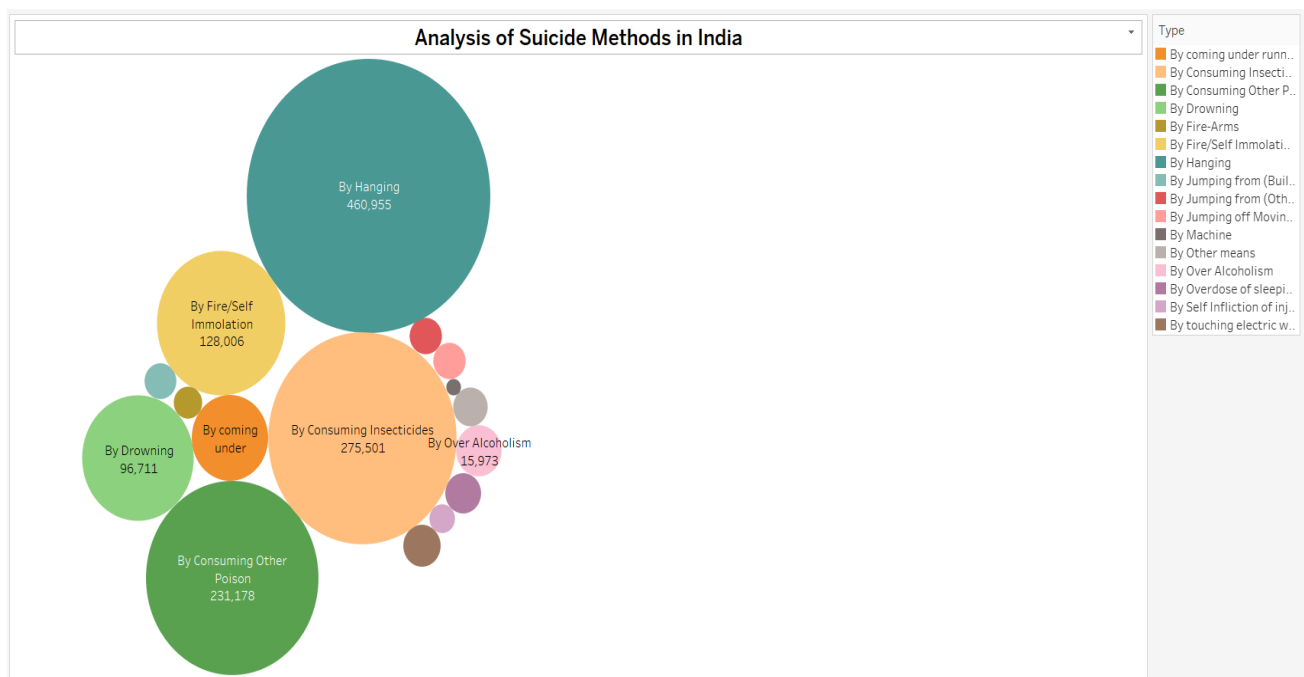
## Results:

The chart compares the total number of suicides across different educational qualifications, further broken down by gender. In this visual, the 'Primary' Education category has the highest counts of suicide incidents, indicating a higher suicide rate among these groups. As the educational qualification increases to Secondary, the total incidents of suicide decrease, suggesting that higher educational attainment might correlate with lower suicide rates. Irrespective of Education more males are committing suicides than females.

## **Bubble Chart**

**Design Rationale:** The bubble chart format is chosen for its ability to represent the volume of data points (suicide cases) through the size of the bubbles, making it easier to compare the prevalence of each method briefly.

**Perceptual Tasks:** Different colors are used to differentiate between methods, aiding in quick visual discrimination. The chart's scale allows for immediate identification of the most and least common methods, with larger bubbles representing higher incidence.



## **Results:**

Hanging is the most prevalent method, followed by consuming insecticides and other poisons, indicating these areas as critical points for intervention. The methods such as over alcoholism and by coming under running vehicles are among the least common, as indicated by the smaller bubble sizes.

## **Conclusion:**

From the above analysis it is evident that the issue of suicide in India is multifaceted and complex, involving various socio-economic, educational, and demographic factors. The analysis highlights significant variances in suicide rates across different states, genders, and over the years from 2001 to 2012. It underscores the importance of addressing mental health, social pressures, and economic challenges as key factors in suicide prevention. The insights gained from this study are crucial for policymakers and healthcare professionals in formulating targeted interventions and support systems to reduce the incidence of suicides in India.

## **References:**

- 1) <https://www.kaggle.com/code/amulya9/data-analysis-of-suicides-in-india>
- 2) <https://jovian.com/abhinav2300/exploratory-data-analysis-and-visualization-of-suicides-in-india>
- 3) <https://www.analyticsvidhya.com/blog/2021/03/exploring-suicide-data-in-india-with-python-code/>
- 4) [https://public.tableau.com/views/Suicidesinindia-acausalanalysis/SuicidesinIndia-acausalanalysis?%3Aembed=y&%3AshowVizHome=no&%3AshowTabs=y&%3Adisplay\\_count=y&%3Adisplay\\_static\\_image=y](https://public.tableau.com/views/Suicidesinindia-acausalanalysis/SuicidesinIndia-acausalanalysis?%3Aembed=y&%3AshowVizHome=no&%3AshowTabs=y&%3Adisplay_count=y&%3Adisplay_static_image=y)